**Having a Vacation in İstanbul, like Living in New York**

Sercan Kaynak

November 29,2020

**1. Introduction**

**1.1. Background**

New York City (NYC), often called simply New York, is the most populous city in the United States. New York is also the most densely populated major city in the United States. Located at the southern tip of the U.S. state of New York, the city is the center of the New York metropolitan area, the largest metropolitan area in the world by urban landmass.

New York has been described as the cultural, financial, and media capital of the world, significantly influencing commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports. Home to headquarters of the United Nations. New York is an important center for international diplomacy.

İstanbul is the most populous city in Turkey and the country’s economic, cultural, and historic center. İstanbul is a transcontinental city in Eurasia, straddling the Bosphorus Strait between the Sea of Marmara and the Black Sea. Istanbul’s commercial and historical center lies on the European side and about a third of its population lives in suburbs on the Asian side of the Bosphrous.

With a total population of around sixteen million residents in İstanbul metropolitan area. İstanbul is one of the world’s largest cities by population, ranking as the world’s fifteenth-largest city and the largest city in Europe. This city is the administrative centre of the İstanbul Metropolitan Municipality.

The city’s biggest attraction is the its historical center, partially listed as UNESCO World Heritage Site, and its cultural and entertainment hub is accross the city’s natural harbor, the Golden Horn,in the Beyoğlu district.

**1.2. Problem**

The Project aims to recommend cheaper places to stay in İstanbul and a chance to find places like people live in their neighborhoods of New York City, if people live in New York want to have a vacation to İstanbul.

Data that might contribute to determining similar places in both cities include geographical coordinates, venues in both cities, and host listings.

**1.2. Interest**

Obviously, people that might have a trip to İstanbul from New York would be very interested in clustered venues of İstanbul for beter vacation experience in İstanbul. Tourism agencies may also be interested.

**2. Data Acquisition and Cleaning**

**2.1. Data Sources**

Geographical coordinates of New York can be found in NYU Spatial Data Repository as GeoJson file.

Geographical coordinates of İstanbul can be built in geojson.io website.

Data of all venues in both İstanbul and New York City can be found in Foursquare with using Foursquare API.

Data Host listings and prices can be found in Airbnb or insideairbnb.com website.

**2.2. Data Cleaning**

All of the data downloaded or scraped from multiple sources. There were some missing values from non-entering neighborhoods,coordinates and prices.

After fixing the problems, I have to make all of data similar columns.

**2.3. Feature Selection**

After data cleaning, 33062 samples and 12 features in all the data. Examining of each features, some of features are redundant features.

After discarding redundant features, I decide to choose four features in both geographical data.

- Borough

- Neighborhood

- Latitude

- Longitude

For host listings data;

- Borough

- Neighborhood

- Latitude

- Longitude

- Property Type

- Room Type

- Price

After all, the features above were selected for the two datasets.

**3. Exploratory Data Analysis**

**3.1. Mapping Neighborhoods for New York City and İstanbul**

We all know boroughs and neighborhoods in the city that we live in. But another cities? We usually take a map and look at the places that we want to go or the places nearby our accomodation places. Cafés, restaurants,shopping centres etc.

First, we have to look at our data for visualizing it.

I look at dataframes for both cities and try to figure it out how many borough and neighborhoods in New York and İstanbul.

In New York City, there is 5 boroughs and 306 neighborhoods.

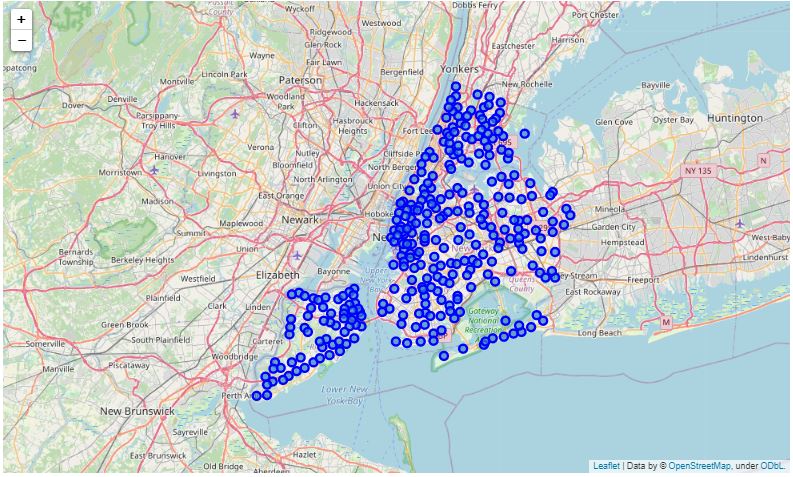
In İstanbul, 39 boroughs and neighborhoods.

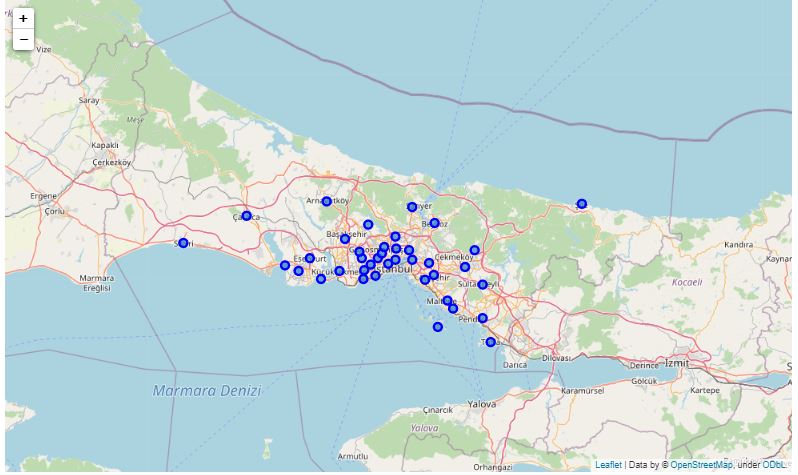
Then to see in that data, I get the geographical coordinates of İstanbul and New York City.

Using Folium library, I visualize both cities boroughs and neighborhoods on maps.

We can see where in boroughs and neighborhoods or nearby or far from coasts.

New York City Map



İstanbul Map

You can see on both map, high density of population usually live nearby coast. Density of population in Manhattan, New York is higher than other boroughs and neighborhoods. If we look at İstanbul, we ‘ll also see the same density there. Density of population throughout Bosphorus is higher than other boroughs and neighborhoods.

**3.2. Exploring Venues in New York City and İstanbul**

We said that the aim of this project is to find similar places between New York City and İstanbul. As second step of EDA, we’ll explore venues in both cities.

After having an account in Foursquare Developers, using Foursquare API, I coded a function to get data about venues. Then I run the function for both cities,to find the venues of cities.

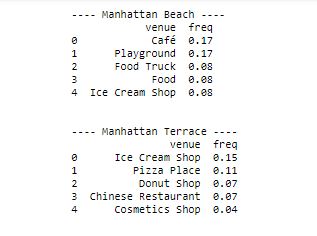
That is a slice from New York venues:



In New York City, there are 10478 venues and from that number of venues,there are 436 unique categories of venues.

After doing one hot encoding of features, I get top 5 venues for each neighborhood.

That is a slice of top 5 venues for each neighborhood:



After determining the top 5 venues, I coded a function to get 10 most common venues for each neighborhoods.

You can see the dataframe that show first five neighborhoods below:



Then I ‘ve done the same process for İstanbul.

I use the same function for venues of İstanbul,using Foursquare API.

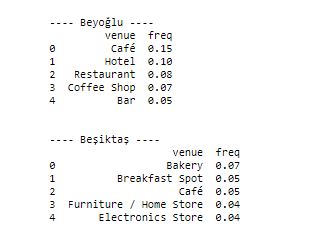
That is a slice of venues:



In İstanbul, there are 2046 venues and from that number of venues,there are 251 unique categories of venues.

After doing one hot encoding of features, I get top 5 venues for each neighborhood.

That is a slice of top 5 venues for each neighborhood:



After determining the top 5 venues, I coded a function to get 10 most common venues for each neighborhoods.

You can see the dataframe that show first five neighborhoods below:



After one hot encoding of features and determining most common venues for New York City and İstanbul, we can find the similarity between two cities.

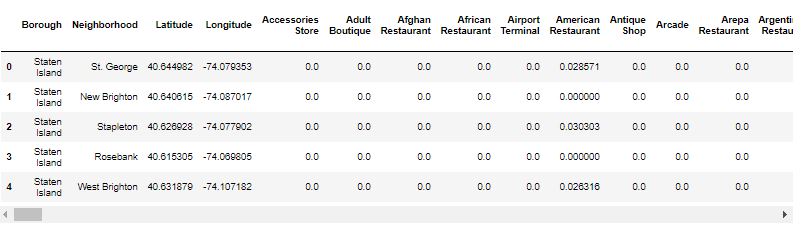
**3.3. Exploratory Data Analysis for similarity between New York City and İstanbul**

First of all, we have to get venue data, do one hot encoding and merge dataframes that shown features; ”Borough”, ”Neighborhood”, ”Latitude”, ”Longitude” and one hot encoding of features.

I coded two functions to get and make a dataframe that I mentioned in the beginning of this section.

After running functions, I take some dataframes.

For New York City (first five indexes) :



For İstanbul (first five indexes) :



After these process, with a for loop, I determine the number of common venue categories in both data and number of different venue categories in both cities.

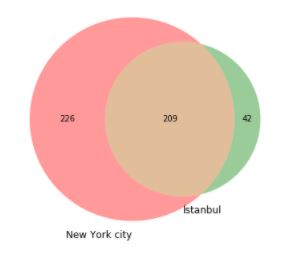
- Number of common venue categories in both data : 209,

- Number of different venue categories in New York City : 226,

- Number of different venue categories in İstanbul : 42

As you can see above, half of New York venues and most of İstanbul venues has common venues.

Now we can visualize that on Venn Diagram.



**3.4. Finding similar venues between New York City and İstanbul**

After EDA for similarity, I coded a for loop for finding common venues and make a list object for analysis.

After that, I converted data to numpy ndarray and find geographical coordinate for one neighborhood from New York to find similar neighborhoods in İstanbul for it.

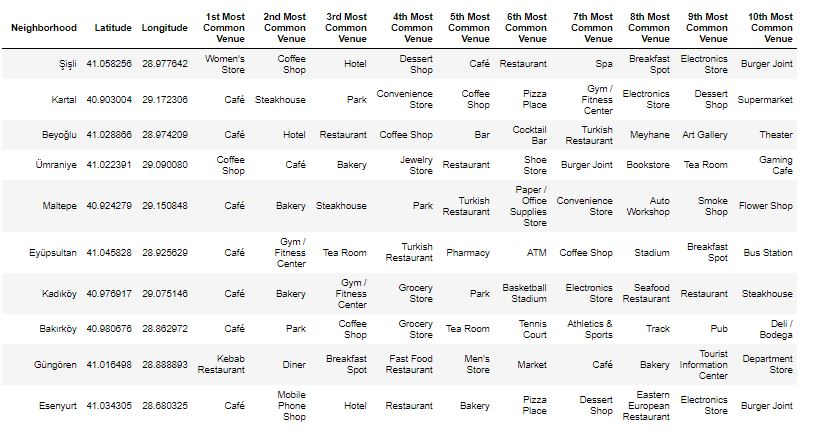
I chose Upper East Side,Manhattan,New York City to find its similar neighborhoods in İstanbul.

After using cosine similarity method, I found 10 most similar neighborhoods in İstanbul.

I found that the most similar neighborhood is Şişli,İstanbul.

Then I merged 10 most similar neighborhoods data with its 10 most common venues.

Also you can see in the dataframe below :

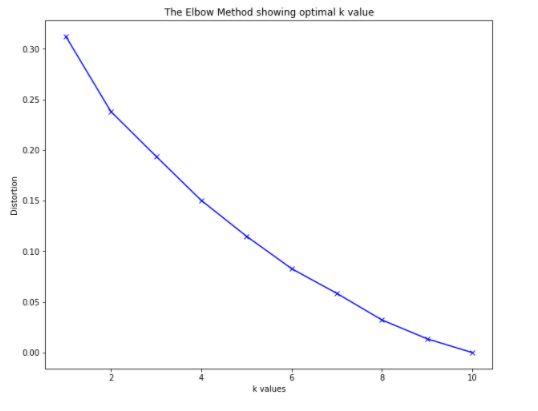


**4. Cluster Modeling**

There are many clustering model in unsupervised machine learning. In this project, we have lots of different data for venues and for clustering, the best algorithm is K-Means Clustering Algorithm.

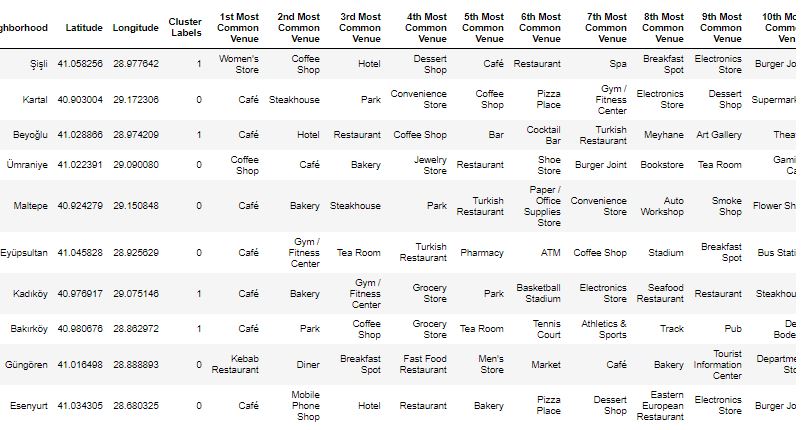
Using K-Means Clustering, I could find how many clusters for similar neighborhoods in İstanbul.

First I chose k=3 but after using elbow method for optimal k, I found optimal k = 2. The reason of lower k value is th data that I used have 10 neighborhood value. Maybe if most similar places was more, k value maybe have would been more than k=2.

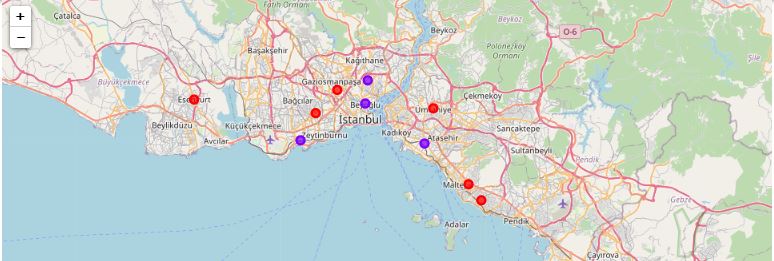


After clustering with K-Means Clustering algorithm, I found all of neighborhoods which cluster labels has.

Then I merged with dataframe that shows neighborhoods, geographical coordinates and ten most common venues.



Then I visualize the clusters on İstanbul map.



After examine the clusters, I named them as:

- Red Dots , Cluster 0 : “Restaurants and Diners Venues”,

- Purple Dots, Cluster 1 : “Cafés & Bars and Multiple Social Venues”

**5. Results**

In the final section of project, I used Airbnb host listing data for cheaper places to stay in similar places of İstanbul.

I used the data from insideairbnb.com website and clean and prepare for the analysis.

For analysis of hosts, I chose neigborhoods in Cluster 1 that including Şişli,İstanbul.

In Cluster 1, I have four neighborhood similar ; Bakırköy, Beyoğlu, Kadıköy, Şişli.

After all process, in four neighborhoods I found number of host for each neighborhoods;

- Bakırköy has 260 hosts,

- Beyoğlu has 5601 hosts,

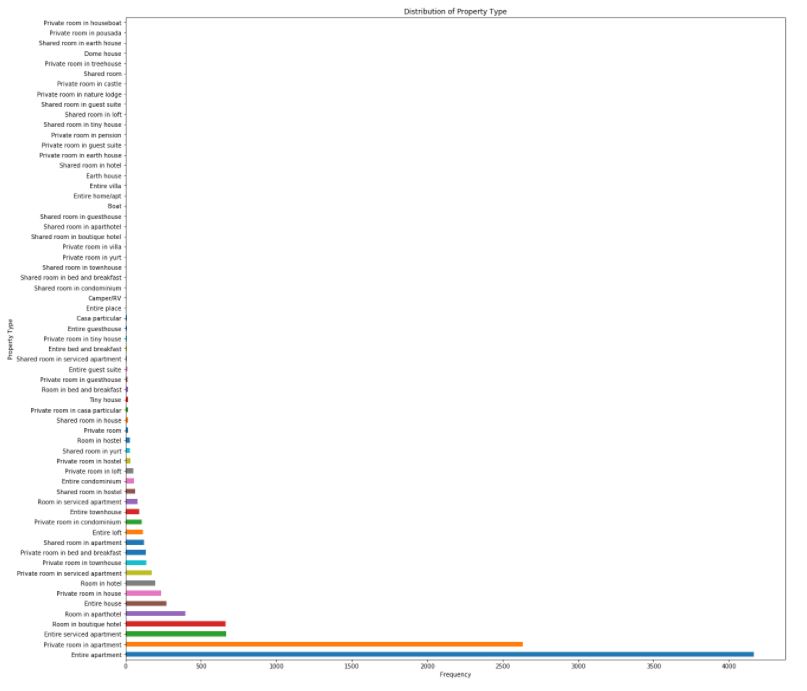
- Kadıköy has 2001 hosts,

- Şişli has 2808 hosts in Airbnb.

After merging all four neighborhoods data, I found the distribution of property types of hosts.

The most frequency of property types of hosts are the hosts that can provide entire apartment and private room in apartment.

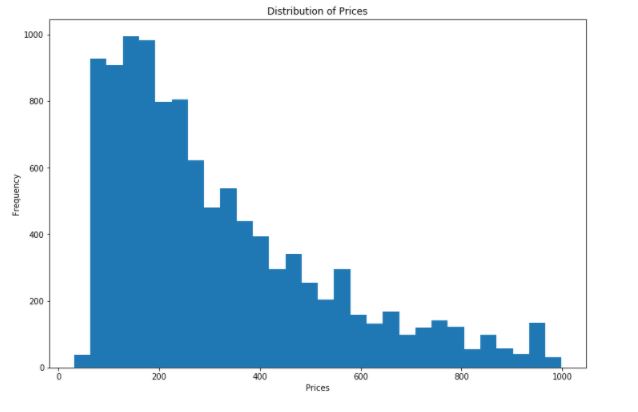
You can see the distribution of property types at horizontal bar chart.



Actually there are lots of kind property types but for a good places to stay, Private room or entire apartment can be chosen.

Then I looked at price distribution and visualize them.

The distribution of prices:



As you can see above, prices to 200 TL is the most given prices from hosts.

After that I could divide prices to five groups.

- Low Prices : prices < 200 TL,

- Low-Middle Prices : 200 TL < prices < 400 TL,

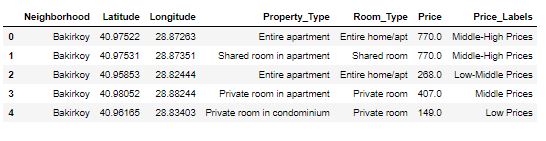
- Middle Prices : 400 TL < prices < 600 TL,

- Middle-High Prices : 600 < prices < 800 TL,

- High Prices : prices > 800 TL.

Then I labeled prices according to these five groups.

There is a slice from it:



Then I made a dataframe all four similar neighborhoods’ most common venues with average prices and price labels.



Finally, I show the distribution of prices on İstanbul Map.



Blue areas show us the lower host prices than others.

We say that the more similar and cheaper place that living like a New Yorker, is Kadıköy.

**6. Conclusion**

As you may know,İstanbul is a big city with multiple venues and areas. Also İstanbul has multiple neigborhoods in each borough. But just 39 district(borough) coordinates were used. For much better analysis and recommendation,the data set can be expanded and the details of places or streets can be also included.

I used KMeans clustering algorithm for clustering similar neighborhoods.And I see on the elbow method optimum k = 2 and used k=2. İstanbul is narrow area and includes various and similar places. If places were more different kind of venues,maybe we can see additional 1 or 2 clusturs in this analysis.

I also used Airbnb host listing data for finding cheaper places to stay in similar neighborhoods of İstanbul that is like in New York City.After I have cleaned the data,I found average prices.

I ended the study by visualizing the host listing data and clusters at similar places on İstanbul Map. Maybe this project can be a pre-idea for tourism agencies.

**7. Future Directions**

As a result, people like experiencing new cultures and traveling a lot. Because of that, people can travel and accomodate cheaper and live in places like their neighborhood along their vacations.

Also investors can provide an information before what kind of venues they invest.